

# Rational truth-avoidance and self-esteem

David Andolfatto *Department of Economics, Simon Fraser University*

Steeve Mongrain *Department of Economics, Simon Fraser University*

Gordon Myers *Department of Economics, Simon Fraser University*

*Abstract.* We assume that people have beliefs about their abilities that generate self-esteem, and that self-esteem is valued intrinsically. Individuals face two choices; one of which strictly dominates the other in a pecuniary sense, but necessarily involves gathering information concerning their ability. We lay out the circumstances under which an individual may find it rational to reject the dominant choice, an act that, in psychology is described as *avoiding the situation*. We then go on to show that the incentive to avoid the truth is increasing in income/wealth and decreasing in self-esteem, the perceived accuracy of one's self-assessment, and the role that luck plays in generating opportunities. JEL classification: D83, D1

*Évitement rationnel de la vérité et estime de soi.* On part des postulats que les gens ont des croyances quant à leurs habiletés et que ces croyances engendrent l'estime de soi, et que l'estime de soi est valorisée en soi. Des individus sont confrontés à une situation de choix entre deux possibilités: l'une de ces possibilités domine strictement l'autre au plan pécuniaire, mais nécessite la cueillette de renseignements concernant les habiletés de la personne. On développe un cadre dans lequel un individu peut trouver rationnel de ne pas choisir le possible dominant; un choix que l'on nomme en psychologie 'éviter la situation'. On montre ensuite que l'incitation à éviter la vérité augmente à proportion que le revenu et la richesse s'accroissent mais diminue à mesure que l'estime de soi, le sens qu'on est exact dans son auto-évaluation, et l'importance qu'on accorde à la chance dans la genèse des opportunités – s'accroissent.

We would like to thank Atsu Amegashie, Stéphane Auray, Dwayne Benjamin, Samuel Danthine, Zane Spindler, and participants at the 2005 World Congress of the Econometric Society, 2004 Journée du CIRPÉE, and the 2003 Queen's University public economics workshop for their comments and suggestions. We are especially indebted to our referees, whose sharp eyes and even sharper criticisms have led to a much improved product. Each author acknowledges the financial support of SSHRC. Email: dandolfa@sfu.ca; mongrain@sfu.ca; gmmymers@sfu.ca

## 1. Introduction

In the standard economic model of individual decision-making, psychological factors play no role in understanding human behaviour. While this approach has proved useful for predicting and interpreting behaviour in a wide variety of contexts, psychologists have apparently identified several phenomena that are not so easily understood under the conventional economic paradigm.

The focus of this paper concerns one such phenomenon, labelled *situation-avoidance* by social psychologists (e.g., see Crocker and Park 2003, 299). To an economist, situation-avoidance might be better labelled *truth-avoidance*, as the phenomenon concerns an action (or inaction) that purposely avoids the acquisition of economically valuable information relating to one's personal characteristics – even when the cost of acquiring such information appears to be zero. Needless to say, such behaviour is logically inconsistent with the basic premises of conventional economic theory. To social psychologists, however, truth-avoidance makes perfect sense, once it is recognized that individuals appear to care directly about how they view themselves; that the acquisition of new information can affect this view; and that human behaviour appears to be governed, at least to some extent, by a concern for protecting or enhancing one's self-esteem. Avoiding the situation (truth-avoidance) is considered the first line of defence among self-esteem management strategies (see Hoyle et al. 1999; Crocker and Park 2003).

Our approach is to take the psychologists' view seriously and at face value. We do this by embedding the underlying psychological assumptions in an otherwise standard economic model. We then investigate and evaluate the logical implications of the hybrid theory. Our economic model is modified by extending the commodity space over which preferences are defined to include an object that reflects an individual's (rational) estimate of his own (unobserved) ability. We label this object *self-esteem*. Individuals are confronted with two choices, one of which strictly dominates the other in a pecuniary sense. The dominant strategy is necessarily associated with gathering information concerning one's (unobserved) ability. The dominated strategy (truth-avoidance) forgoes an obvious pecuniary gain and reveals no information. We restrict belief-formation to be rational (in the sense of respecting Bayes' rule). As such, individual actions (or inactions) influence the evolution of one's self-assessment over time. For individuals who do not value self-esteem, the evolution of this self-assessment is immaterial (the dominant strategy is always preferred). However, under specific circumstances, truth-avoidance can be consistent with rational behaviour.

The simple fact that people care about self-esteem (or possess 'ego-utility,' to borrow a term from Köszegi 2006) does not, in itself, imply anything about behaviour. In other words, simply 'sticking self-esteem into the utility function' does not necessarily place restrictions on behaviour. As is well known, strict concavity of the self-esteem utility function is needed. That is, 'good news' that would lead one to revise upward one's estimate of one's ability must be valued less than 'bad news' that would lead to an equivalent downward revision. Hence,

concavity of utility over self-esteem implies a form of ‘information aversion.’ Naturally, avoiding the truth further requires that individuals believe that they will learn something about themselves from facing the new information. This requires that individuals lack ‘confidence’ in the accuracy of their own self-assessment and that individuals perceive that future opportunities are driven at least partly by skill (relative to luck). These results are by now generally well known in the economics literature.

The focus of our modelling, however, is to characterize the nature of individuals who are likely to display a propensity for truth-avoidance. We find that, *ceteris paribus*, high-income/wealth individuals are more likely to avoid the truth and that individuals with high self-esteem are less likely to avoid the truth. In other words, truth-avoiders tend to be those with incomes that are high relative to the self-assessment of their own ability. To the extent that income and self-esteem are not perfectly correlated within a population, the phenomenon of truth-avoidance is therefore likely to present among all sorts of individuals. We also show that the propensity for truth avoidance is decreasing in an individual’s ‘confidence’ in the accuracy of self-assessment and the extent to which the individual perceives future opportunities to be driven by luck. These are all potentially testable implications, given the type of data that are commonly produced by questionnaires and experiments conducted by psychologists.

These results may be of interest to both economists and social psychologists. For the economist, our findings suggest that the traditional practice of ignoring psychological factors like self-esteem may be justified in some circumstances but not in others. These circumstances include environments where individuals are likely to be sufficiently confident in the accuracy of their self-assessment (which is not the same thing as saying that they are necessarily accurate in their self-assessment); or when they do not display ‘information aversion.’ For the psychologist, our findings identify various individual characteristics that are likely to render individuals more or less prone to avoiding the truth. Among other things, the theory developed here may be useful in guiding experimental design.

Our paper fits within a growing body of theoretical work designed to explain what, on the surface at least, appears to be ‘anomalous’ economic behaviour. One strand of this literature simply assumes that individuals are prone to making cognitive mistakes (e.g., see Rabin and Schrag 1999; Gervais and Odean 1999). Another strand of the literature, exemplified by the recent work of Benabou and Tirole (2002a,b), models the manipulation of self-image as a strategic game played among time-dated personalities.

Our own approach is most closely related to Köszegi (2001, 2006) and Weinberg (2004), who, like ourselves, model self-esteem as ego-utility. Our paper differs from these primarily in focus and the particular questions addressed. Köszegi (2006) and Weinberg (2004) are primarily concerned with explaining how people may rationally become overconfident (something that we do not address). Köszegi (2001), on the other hand, explains why it may be rational for people to avoid reviewing new information concerning past decisions and why it

may be rational to procrastinate in making decisions when there is no pecuniary gain from doing so. We view our paper as complementary to this literature, as it simplifies along some dimensions, but delves deeper along others.

## 2. Basic model

Consider an economy with people who have preferences defined over lotteries of consumption  $c \in \mathbb{R}$ . These preferences are represented by an expected utility function  $E[u(c)]$ , where  $E$  denotes an expectations operator and  $u'' \leq 0 < u'$ . Each person has an initial endowment  $(w, z) \in \mathbb{R}_+^2$ , which is distributed in some arbitrary manner across the population. The parameter  $w$  represents the return associated with some economic opportunity (the quality of a job, investment, mate, etc.), while  $z$  represents non-labour income.

Each individual may take one of two actions, which we denote  $I \in \{0, 1\}$ . The action  $I = 0$  corresponds to consuming one's initial endowment, so that  $c = w + z$ . The action  $I = 1$  corresponds to an act that may potentially improve one's circumstance. We model this potential improvement as a new opportunity, whose value  $w'$  is determined by the random process:

$$w' = a + e, \tag{1}$$

where  $a$  represents an endowed 'ability' and  $e$  represents 'luck.' Assume that ability is distributed across the population in a Gaussian manner. Furthermore, assume that each individual faces an *i.i.d.*  $e \sim N(0, \sigma^2)$ .<sup>1</sup>

Assume that the action  $I = 1$  entails no pecuniary cost. Assume, further, that one always retains the option of discarding the new opportunity  $w'$  in favour of the old  $w$ , so that  $c = \max\{w + z, w' + z\}$ . One interpretation of this model is that  $I = 1$  represents a job-search activity (with perfect recall), with a wage offer that depends in part on skill and in part on match-quality. The choice  $I = 0$  would in this case represent declining the search option. In the language of social psychology, we want to think of  $I = 0$  as corresponding to 'situation-avoidance.' Given that there is absolutely no cost to 'facing the situation,' the only rational choice here would be  $I = 1$ .

### 2.1. Information and beliefs

In general, an individual may not know with certainty his or her own true ability level  $a$ . In this case, we assume that individuals are Bayesian. In the present context, since  $(a, e)$  are distributed joint-normally, Bayes' rule corresponds to the Kalman filter (see Ljungqvist and Sargent 2000, 65–71).

<sup>1</sup> Implicitly then, we allow for negative consumption. However, one could guarantee positive consumption by assuming instead that  $c = \exp(w) + z$ . As nothing in our analysis hinges on this matter, we allow for negative consumption only to simplify notation.

That is, imagine that each person begins with a prior  $(b, \Sigma)$ , so that one's ability is perceived to be distributed normally with mean  $b$  and variance  $\Sigma = E[a - b]^2$ . Since  $b$  represents a person's self-assessment of his own ability, we refer to  $b$  as 'self-esteem.' Note that  $\Sigma$  is a parameter that describes an individual's 'confidence' in his self-assessment. In particular,  $\Sigma^{-1}$  is referred to as the *precision of the estimate* ( $b$ ), so that  $\Sigma^{-1} = \infty$  represents the case of an individual who is supremely confident in his self-assessment (which is not to say that the self-assessment is necessarily correct).

Now, conditional on  $I = 1$ , an individual generates a new opportunity  $w' = a + e$ . Not knowing one's true ability, however, implies that an individual faces a signal-extraction problem. Given  $b$  and the new information associated with  $w'$ , an individual will update his self-assessment  $b' = E[a \mid b, w']$  according to

$$b' = (1 - k)b + kw', \quad (2)$$

where

$$k = \frac{\Sigma}{\Sigma + \sigma^2}. \quad (3)$$

In addition, the perceived precision of one's self-assessment evolves according to

$$\Sigma' = \frac{\sigma^2}{\Sigma + \sigma^2} \Sigma. \quad (4)$$

Of course, in the case of  $I = 0$  (truth-avoidance), no new information is gathered, so that

$$\begin{aligned} b' &= b \\ \Sigma' &= \Sigma. \end{aligned} \quad (5)$$

Thus, (2) asserts that  $b'$  is given by a convex combination of one's prior  $b$  and new information  $w'$  (in the case for which  $I = 1$ ). The Kalman-gain variable  $k$  determines how much weight is to be placed on the latter two objects. For a given  $\Sigma$ , we see from (3) that  $k$  is decreasing in the 'noise' term  $\sigma$ . That is, if the value of a new opportunity is determined primarily by luck rather than ability (i.e., a large  $\sigma$ ), then any optimal reassessment of ability should largely ignore new information and rely more heavily on prior beliefs.

As well, note that for a given  $\sigma$ , (3) also reveals that  $k$  is a decreasing function of  $\Sigma^{-1}$ . Recall that  $\Sigma^{-1}$  measures the (perceived) precision of one's current estimate of ability. As  $\Sigma \rightarrow 0$ , one becomes increasingly 'confident' in one's self-assessment, so that  $k \rightarrow 0$  (it is optimal to ignore noisy information and rely more heavily on prior beliefs). Equation (4) describes how the precision of one's self-assessment evolves.<sup>2</sup>

2 Extending this model to an infinite horizon, equation (4) implies that infinitely-lived individuals could potentially learn their true ability (i.e.,  $\Sigma_t \rightarrow 0$  as  $t \rightarrow \infty$ ).

Modifying the information structure in this manner affects how individuals form expectations, but otherwise does not affect behaviour (it still remains optimal to choose  $I = 1$ , regardless of one's initial condition as summarized by the triplet  $(w, b, \Sigma)$ ). To show this formally, let  $F(w', b)$  denote the cumulative distribution function for  $w'$  conditional on  $b$ . Then the expected utility payoff associated with  $I = 1$  is given by

$$E \max\{u(w + z), u(w' + z)\} = \int_w u(w' + z)F(dw', b) + F(w, b)u(w + z).$$

This obviously dominates the utility payoff associated with avoiding the situation; i.e.  $E \max\{u(w + z), u(w' + z)\} \geq u(w + z)$ .<sup>3</sup>

### 3. A model of self-esteem

In the model above, individuals are endowed with some prior  $b$  that measures their assessment of their own (unobserved) ability. Individuals are also endowed with a prior view  $\Sigma$  that measures the perceived precision of their self-assessment. In the basic model outlined above, neither of these objects plays a role in determining behaviour. Such a view is contrary to social psychology, where the conventional wisdom is that we can not understand individual behaviour without first having an understanding of self-esteem (i.e., see Leary and Tangney 2003).

We model the psychologist's view here by extending the commodity space to include lotteries over the *posterior* belief  $b'$ , so that preferences can be represented by an expected utility function:

$$E[u(c) + \lambda v(b')]. \quad (6)$$

The parameter  $\lambda \geq 0$  in (6) simply indexes the degree to which a person cares about his self-esteem. The model presented earlier is just the special case in which  $\lambda = 0$ . Following Köszegi (2001), we assume that  $v$  is strictly increasing and weakly concave. Strict concavity of  $v$  implies that the person displays a form of 'information aversion.' That is, good news that would lead one to revise upward one's estimate of one's ability is valued less than bad news that would lead to an equivalent downward revision. Because we do not want our results to hinge on hard-to-interpret third-derivative properties of  $v$ , we assume for simplicity that  $v$  takes a quadratic form; that is,

$$v(b) = \alpha b - 0.5\beta b^2, \quad (7)$$

where  $\alpha, \beta \geq 0$  are parameters.<sup>4</sup>

<sup>3</sup> As one of our referees has pointed out, it is crucial here that  $I = 0$  does not depend on ability, while the return to option  $I = 1$  does.

<sup>4</sup> We will also restrict attention to cases where  $b < \alpha/\beta$ , so that  $v'(b) > 0$  always.

### 3.1. Optimal decision-making

Consider an individual described by the list of parameters  $(w, b, \Sigma, \lambda, \alpha, \beta, \sigma)$  describing preferences, technology, and information. This person must make a choice  $I \in \{0, 1\}$  to maximize (6) subject to  $c = \max\{w' + z, w + z\}$  and subject to the rational updating of beliefs (2)–(5).

The utility payoff associated with  $I = 0$  is given by

$$V_0(w, z, b) = u(w + z) + \lambda v(b). \quad (8)$$

The expected utility payoff associated with  $I = 1$  is given by

$$\begin{aligned} V_1(w, z, b) &= E \max\{u(w' + z) + \lambda v(b'), u(w + z) + \lambda v(b')\} \\ &= \int \max\{u(w' + z), u(w + z)\} F(dw', b) + \lambda \int v(b') F(dw', b) \\ &= \int_w u(w' + z) F(dw', b) + u(w + z) F(w, b) + \lambda \int v(b') F(dw', b), \end{aligned} \quad (9)$$

where  $b'$  satisfies (2).

Observe that, while  $I = 0$  removes the attractive option of potentially upgrading the value of one's opportunity, it has the benefit of preserving one's self-esteem (since no information is gathered that would necessarily lead one to update one's belief). For obvious reasons, we label such an action *truth-avoidance*. Conversely, while  $I = 1$  represents an expected pecuniary gain; such an action exposes a person to 'self-esteem risk.' To the extent that an individual cares about self-esteem, the economically rational (i.e., utility-maximizing) choice is no longer obvious; that is, depending on parameters, it is possible that  $V_1(w, z, b) \geq V_0(w, z, b)$ .

Define the following two terms:

$$\Pi(w, z, b) \equiv \int_w u(w' + z) F(dw', b) + u(w + z) F(w, b) - u(w + z) \quad (10)$$

$$\Delta(b) \equiv \lambda \left[ v(b) - \int v(b') F(dw', b) \right]. \quad (11)$$

Here,  $\Pi$  represents the net 'pecuniary' gain from gathering information and  $\Delta$  represents the net 'non-pecuniary' benefit associated with preserving one's self-esteem. Note that  $V_1(w, z, b) - V_0(w, z, b) \equiv \Pi(w, z, b) - \Delta(b)$ .

Observe that the net pecuniary gain from gathering information is monotonically decreasing in  $w$ ; that is,

$$\frac{\partial \Pi}{\partial w} = u'(w + z) [F(w, b) - 1] < 0. \quad (12)$$

Intuitively, the closer one is to the top of the wage distribution, the less likely one is to draw a new opportunity that dominates the one in hand. Note that this result is independent of the curvature properties of  $u$ ; that is, it continues to hold when  $u'' = 0$  (or  $u' = \kappa > 0$  constant). For a given  $\Delta$ , then, one can characterize a reservation wage  $w_R$  satisfying

$$\Pi(w_R, z, b) = \Delta(b). \quad (13)$$

The optimal strategy may therefore be expressed as

$$I^* = \begin{cases} 0 & \text{if } w_R(z, b) < w \leq \infty \\ 1 & \text{otherwise} \end{cases}. \quad (14)$$

Notice that if  $\Delta(b) \equiv 0$ , as in the conventional model studied earlier, then  $w_R = \infty$ . In other words, it strictly pays to ‘face the situation’ for anyone with an endowed opportunity  $w$  below the upper bound of the wage distribution when self-esteem is not a factor in the decision-making process. Of course, if  $\Delta > 0$  is a possibility, then  $w_R < \infty$ , so that, in general, there will be circumstances in which truth-avoidance constitutes a rational choice. These circumstances are now described.

**PROPOSITION 1.** *For  $v(b)$  satisfying (7),  $\Delta = 0.5\lambda\beta\Sigma^2(\sigma^2 + \Sigma)^{-1}$ ; so that  $\Delta > 0$  if and only if  $\lambda > 0$ ,  $\beta > 0$ ,  $\Sigma > 0$ , and  $\sigma^2 < \infty$ .*

The proof of this proposition is provided in the appendix. The logic for the necessary conditions for truth-avoidance,  $\lambda > 0$ ,  $\beta > 0$ , is obvious. Naturally, an additional requirement is that the individual believes that he will learn about himself from facing the new information. This requires that an individual lack ‘confidence’ in the accuracy of his self-assessment  $\Sigma > 0$  and that an individual perceive that future opportunities are driven at least partly by skill (relative to luck)  $\sigma^2 < \infty$ . These results are by now generally well known in the economics literature.

#### 4. Characteristics of truth-avoiders

In this section, we turn to the focus of the paper, the attempt to flesh out some of the characteristics of truth-avoiders. Our approach here is to consider a population of individuals who are identical in every respect, except along one particular dimension in the parameter vector  $(w, z, b, \Sigma, \lambda, \alpha, \beta, \sigma)$ . Of course, the analysis that follows assumes that  $\Delta > 0$ .



#### 4.1. Value of endowments

Imagine a population of individuals who differ only in terms of the value of their current economic opportunity  $w$ . It follows directly from the optimal strategy described by (14):

**PROPOSITION 2.** (*Ceteris paribus*) *Truth-avoiders will be concentrated among those who are currently endowed with relatively good economic opportunities.*

At first, this result may sound surprising, but, in fact, the intuition is simple and follows directly from the fact that  $\Pi$  is monotonically decreasing in  $w$ . In particular, note that while there is no pecuniary cost to gathering information, the upside from doing so is relatively small for those already close to the top. Likewise, for those near the bottom, the upside potential is relatively large. Thus, for a given  $(b, \Delta)$ , the former group has a stronger incentive to avoid the truth. Note that this result does not hinge on the curvature properties of  $u$  (in particular, the result continues to hold even in the case  $u'' = 0$ ).

Let us now imagine a population of individuals who differ solely in terms of their endowed ‘wealth’ as measured by  $z$  (non-labour income).<sup>5</sup> The relevant comparative static here is obtained from (13):

$$\frac{dw_R}{dz} = -\frac{\partial \Pi / \partial z}{\partial \Pi / \partial w},$$

where

$$\frac{\partial \Pi}{\partial z} = \int_w u'(w' + z)F(dw', b) - u'(w + z)[1 - F(w, b)].$$

The fact that  $\partial \Pi / \partial w < 0$  has been established in (12). The first thing we can establish is that if individuals are risk neutral in the sense that  $u'' = 0$  (or  $u' = \kappa > 0$  a constant), then

$$\frac{\partial \Pi}{\partial z} = \kappa [1 - F(w, b)] - \kappa [1 - F(w, b)] = 0.$$

In other words, the propensity for truth-avoidance is unrelated to wealth when individuals are risk neutral. Let us now suppose  $u'' < 0$  and define  $\kappa \equiv u'(w + z) > 0$ . It then follows that

$$\int_w u'(w' + z)F(dw', b) < \kappa [1 - F(w, b)],$$

so that  $\partial \Pi / \partial z < 0$ , which implies  $dw_R / dz < 0$ . One can therefore establish

<sup>5</sup> We thank a referee for suggesting this exercise.

PROPOSITION 3. (*Ceteris paribus*) If  $u'' < 0$ , truth-avoiders will be concentrated among the rich.

The intuition for this result is also rather straightforward. Given the concavity of  $u$ , higher levels of wealth imply that the marginal utility associated with searching for better options is lower. In other words, there is a sense in which the wealthy can better afford to take actions (or inactions) that protect their self-esteem.

#### 4.2. The level of self-esteem

Imagine now a population that differs only in prior self-assessment  $b$ . The relevant comparative static is again obtained from (13):

$$\frac{dw_R}{db} = -\frac{\partial \Pi / \partial b}{\partial \Pi / \partial w}.$$

The fact that  $\partial \Pi / \partial w < 0$  has been established in (12). As  $F(w, b)$  denotes the distribution of wage opportunities conditional on a (perceived) mean  $b$ , one would expect that the net pecuniary gain to gathering information is increasing in  $b$ , that is, that  $\partial \Pi / \partial b > 0$ . One can establish this formally by noting that  $F(w, b_l) > F(w, b_h)$  for  $b_l < b_h$  (i.e., the former conditional distribution stochastically dominates the latter). It follows then that  $\Pi(w, z, b_h) > \Pi(w, z, b_l)$  and  $\partial \Pi / \partial b > 0$  then emerges as one takes  $b_h \rightarrow b_l$ . As a consequence, it follows that  $dw_R/db > 0$ .

In other words, a higher level of self-esteem increases one's reservation opportunity level, thereby reducing the range of values of  $w$  for which it makes sense to engage in truth-avoidance. Someone with high self-esteem finds it more costly to engage in truth-avoidance simply because he expects a better outcome by accepting the signal.<sup>6</sup> From this result, we have the following proposition:

PROPOSITION 4. (*Ceteris paribus*) Truth-avoiders will consist of those who are currently endowed with relatively low self-esteem.

Taken together, propositions 2 and 4 suggest that the phenomenon of truth-avoidance is likely to be concentrated among those individuals who are in some sense 'doing well' relative to their self-assessment of ability (i.e., a high  $w/b$  ratio). Thus, our theory suggests that the phenomenon of truth-avoidance is likely to be found among individuals throughout the income distribution. What matters in our model is not the level of income or self-esteem but rather their relative magnitudes. This is potentially a testable implication.

<sup>6</sup> It should be noted that this result is in part due to the fact in our model, the value of self-esteem is independent of an individual's prior. This, does not need to be the case. For example, individuals with higher initial self-esteem may have intrinsically more to lose from learning the truth. For a more thorough discussion on the topic, please see Benabou and Tirole (2002).

### 4.3. Preferences

Obviously, the propensity to avoid the truth is increasing in  $\lambda$  and  $\beta$ . This non-surprising result follows directly from (14). It is worth repeating that while psychologists appear to emphasize the parameter  $\lambda$ , the growing economics literature on the subject shows that some degree of ‘information-aversion’ is required as well. There are experiments that attempt to measure the affective or emotional response of individuals to given self-esteem-damaging events (Hoyle et al. 1999, 87). The evidence is that individuals with low self-esteem have a stronger affective response to a given self-esteem-damaging failure.

### 4.4. Confidence in one’s self-assessment

Recall that  $\Sigma^{-1}$  represents the perceived accuracy of one’s self-assessment. From proposition 1, it is easy to verify that  $\Delta$  is an increasing function of  $\Sigma$  (i.e., a decreasing function of  $\Sigma^{-1}$ ); this operates through the effect that  $\Sigma$  has on  $k$  in (3). In words, what this means is that among otherwise similar people, those who are confident in the accuracy of their self-assessment are less likely to engage in truth-avoidance.

**PROPOSITION 5.** *The propensity to avoid the truth is decreasing in the perceived accuracy of one’s self-assessment.*

Proposition 5 suggests that the following is possible. Consider two people  $i = 1, 2$ , who are identical in every way except for differences in  $(b_i, \Sigma_i)$ . Suppose that  $b_1 > b_2$  and  $\Sigma_1 > \Sigma_2 = 0$ . That is, while the type-1 person has relatively high self-esteem, this high self-assessment is associated with a degree of uncertainty. The type-2 person, on the other hand, is absolutely confident in his/her self-assessment. Our theory asserts that the low-esteem type-2 person will not practice truth-avoidance; while the high-esteem type-1 person may.

### 4.5. Noisiness of information

**PROPOSITION 6.** *The propensity to avoid the truth decreases with the noisiness of the new signal  $\sigma^2$ .*

$$\frac{dw_R}{d\sigma^2} = \frac{\partial \Delta / \partial \sigma^2 - \partial \Pi / \partial \sigma^2}{\partial \Pi / \partial w}.$$

Recall that  $\partial \Pi / \partial w < 0$ . Also recall that  $\sigma^2$  measures a person’s perception of the relative role that luck plays in determining the value of future opportunities. From  $\Delta$  in proposition 1, we have  $\partial \Delta / \partial \sigma^2 < 0$ . The intuition is that the greater the perceived role of luck (noisy signals), the less costly will it be to expose oneself to information-gathering activities that may damage self-esteem. From (10) the effect of  $\sigma$  on  $\Pi$  simply reinforces the effect on  $\Delta$  or  $\partial \Pi / \partial \sigma^2 > 0$ . The intuition

for this is simple. Observe that an increase in  $\sigma$  represents a mean-preserving spread of the (normal) distribution  $F$ . Since individuals have complete recall, a mean-preserving spread in  $F$  serves to increase the upside potential of gathering information, without altering the downside risk. Thus,  $dw_R/d\sigma^2 > 0$  and an increase in  $\sigma^2$  leads to a decline in the propensity to avoid the truth.

Propositions 5 and 6 are interesting because they suggest that increases in ‘uncertainty’ along different dimensions can have different effects on the propensity for truth-avoidance. That is, increased uncertainty about how past events have influenced the accuracy of one’s current self-assessment (an increase in  $\Sigma$ ) serves to increase truth-avoidance. In contrast, increased uncertainty about the value of future opportunities serves to decrease truth-avoidance.

## 5. Conclusion

It is known that several different forces need to be in operation *simultaneously* if avoiding the truth is likely to manifest itself as observed behaviour. For rational truth-avoidance to arise, people obviously have to care about self-esteem, but in a particular way (they must be averse to self-esteem risk). Further, the individual must believe that facing the situation will be informative. This implies that individuals must be less than fully confident in the accuracy of their self-assessment, and, on top of this, individuals must believe that luck plays only some role in determining their opportunities. Our further analysis suggests that the phenomenon of truth-avoidance is likely to be concentrated among those individuals who are in some sense ‘doing well’ relative to their self-assessment of ability. We also show that the propensity for truth avoidance is decreasing in an individual’s ‘confidence’ in the accuracy of self-assessment and the extent to which the individual perceives future opportunities to be driven by luck. All these are dimensions that are potentially measurable with well-designed experiments or other methods.

For the economist, many types of economic behaviour probably remain plausibly interpretable within the context of theories that abstract from self-esteem issues. But there may be some (perhaps even a great number of) phenomena for which self-esteem issues may play a prominent role. One important example may concern the question of the optimal design of social insurance mechanisms. Casual empiricism suggests that when a member of society ‘hits bottom’ (job loss, divorce, poverty, etc.), low self-esteem becomes an issue. It becomes an issue because of the possibility that low self-esteem can be self-perpetuating (e.g., by abstaining from job-search when the costs of doing so are low). In the context of the theory developed above (extended to many periods), it is possible for a string of unlucky events to drive one’s self-assessment far below one’s true ability, ultimately culminating in a state of perpetual truth-avoidance. How should policy be designed to deal with such a scenario? We believe that this is a promising area of future research.

## Appendix

### *Proof of proposition 1*

$$\Delta \equiv \lambda \{v(b) - E[v(b') \mid b, \Sigma]\},$$

where  $E[\bullet \mid b, \Sigma]$  denotes the expectation conditional on  $b$  and  $\Sigma$ . Using the quadratic form for  $v$ , we have

$$\begin{aligned} v(b) &= \alpha b - (1/2)\beta b^2 \\ v(b') &= \alpha b' - (1/2)\beta (b')^2. \end{aligned}$$

Observe that

$$\begin{aligned} E[v(b') \mid b, \Sigma] &= \alpha E[b' \mid b, \Sigma] - (1/2)\beta E[(b')^2 \mid b, \Sigma] \\ &= \alpha b - (1/2)\beta E[(b')^2 \mid b, \Sigma], \end{aligned}$$

as  $E[b' \mid b, \Sigma] = (1 - k)b + kE[w' \mid b, \Sigma] = b$ . Combining what we have so far:

$$\Delta = \lambda(1/2)\beta \{E[(b')^2 \mid b, \Sigma] - b^2\}.$$

We can now expand the term  $E[(b')^2 \mid b, \Sigma]$ ; that is,

$$\begin{aligned} E[(b')^2 \mid b, \Sigma] &= E[(kw' + (1 - k)b]^2 \mid b, \Sigma] \\ &= E[k^2(w')^2 + 2k(1 - k)bw' + (1 - k)^2b^2 \mid b, \Sigma] \\ &= k^2 E[(w')^2 \mid b, \Sigma] + 2k(1 - k)bE[w' \mid b, \Sigma] + (1 - k)^2b^2 \\ &= k^2 E[(w')^2 \mid b, \Sigma] + (1 - k)b^2[2k + (1 - k)] \\ &= k^2 E[(w')^2 \mid b, \Sigma] + (1 - k^2)b^2. \end{aligned}$$

We still need to expand the term  $E[(w')^2 \mid b, \Sigma]$ ; that is,

$$\begin{aligned} E[(w')^2 \mid b, \Sigma] &= E[(a + e')^2 \mid b, \Sigma]; \\ &= E[a^2 \mid b, \Sigma] + 2E[ae' \mid b, \Sigma] + \sigma^2. \end{aligned}$$

Given that  $a$  and  $e'$  are independent, we know that  $E[ae' \mid b, \Sigma] = 0$ . Moreover, we also know that  $E[a^2 \mid b, \Sigma] = \Sigma + b^2$ . Consequently, we get

$$E[(w')^2 \mid b, \Sigma] = b^2 + \Sigma + \sigma^2.$$

Thus, we are left with

$$\begin{aligned} \Delta &= \lambda(1/2)\beta[(1 - k^2)b^2 + k^2b^2k^2(\sigma^2 + \Sigma) - b^2] \\ &= \lambda(1/2)\beta \frac{\Sigma^2}{\sigma^2 + \Sigma}. \end{aligned}$$

## References

- Benabou, R., and J. Tirole (2002a) 'Self-confidence and personal motivation,' *Quarterly Journal of Economics* 117, 871–915
- (2002b) 'Intrinsic and extrinsic motivation,' *Review of Economic Studies* 70, 489–520
- Berglas, S., and E.E. Jones (1978) 'Drug choice as a self-handicapping strategy in response to noncontingent success,' *Journal of Personality and Social Psychology* 36, 405–17
- Crocker, J., and L.E. Park (2003) 'Seeking self-esteem: construction, maintenance, and protection of self-worth,' in *Handbook of Self and Identity*, ed. M.R. Leary and J.P. Tangney (New York: Guilford Press)
- Gervais, S., and T. Odean (1999) 'Learning to be overconfident,' manuscript, University of California, Davis
- Hoyle, R.H., M.H. Kernis, M.R. Leary, and M.W. Baldwin (1999) *Selfhood* (Boulder, CO: Westview Press)
- Köszegi, B. (2001) 'Ego utility and information acquisition,' manuscript, University of California, Berkeley
- (2006) 'Ego utility, overconfidence, and task choice,' forthcoming, *Journal of the European Economic Association*
- Leary, M.R., and J.P. Tangney (2003) 'Preface' in *Handbook of Self and Identity*, ed. M.R. Leary and J.P. Tangney (New York: Guilford Press)
- Ljungqvist, L., and T.J. Sargent (2000) *Recursive Macroeconomic Theory* (Cambridge, MA: MIT Press)
- Rabin, M., and J. Schrag (1999) 'First impressions matter: a model of confirmatory bias,' *Quarterly Journal of Economics* 114, 37–82
- Shepperd, J.A., and R.M. Arkin (1989) 'Determinants of self-handicapping: task importance and the effects of preexisting handicaps on self-generated handicaps,' *Personality and Social Psychology Bulletin* 15, 101–12
- Weinberg, B. (2004) 'A model of overconfidence,' manuscript, Ohio State University